DOCUMENT RESUME

ED 064 354 TM 001 522

AUTHOR Proger, Barton B.; And Others
TITLE Adapting Criterion-Referenced Measurement to Individualization of Instruction for Handicapped Children: Some Issues and a First Attempt.
PUB DATE Mar 72
NOTE 24p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Tests; *Criterion Referenced Tests; Evaluation Techniques; Handicapped Children; *Individual Tests; *Measurement Instruments; *Mentally Handicapped; Test Construction
IDENTIFIERS IAMS; *Individual Achievement Monitoring System

ABSTRACT

Criterion-referenced measurement (CRM) has received increasing attention in regular education. However, it is in education for handicapped children that CRM's flexibility for individualization of both instruction and evaluation become even more fully realized. Research is described on one of the first CRM systems (Individual Achievement Monitoring System: IAMS) ever devised exclusively for the handicapped and designed for widespread implementation. Methodological problems are discussed, such as inappropriateness of item sampling, difficulties in retention testing, and determination of adequate criterion levels of mastery for handicapped children. Flexibility of research findings based on CRM is also examined. (Author)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 064354

ADAPTING CRITERION-REFERENCED MEASUREMENT TO INDIVIDUALIZATION OF INSTRUCTION FOR HANDICAPPED CHILDREN: SOME ISSUES AND A FIRST ATTEMPT[1]

by

Barton B. Proger, Lester Mann, Robert M. Burger, Lawrence H. Cross

Pennsylvania Resources and Information Center for Special Education

King of Prussia, Pennsylvania[2]

March, 1972

TM 001 522

A paper presented at the 1972 annual convention of the American Educational Research Association, Division D (Measurement and Research Methodology), Chicago, Illinois, April, 1972

FILMED FROM BEST AVAILABLE COPY

(Abstract)

Criterion-referenced measurement (CRM) has received increasing attention in regular education. However, it is in education for handicapped children that CRM's flexibility for individualization of both instruction and evaluation become even more fully realized. Research is described on one of the first CRM systems (Individual Achievement Monitoring System: IAMS) ever devised exclusively for the handicapped and designed for widespread implementation. Methodological problems are discussed, such as inappropriateness of item sampling, difficulties in retention testing, and determination of adequate criterion levels of mastery for handicapped children. Flexibility of research findings based on CRM is also examined.

INTRODUCTION

There is a great deal of interest in criterion-referenced measurement (CRM) at present. This year's convention alone reflects this ever-increasing attention. Several advantages to CRM seem to be fostering its growth. First, some like the idea that CRM, unlike norm-referenced measurement (NRM), does not force the test interpreter into making sometimes inappropriate group comparisons. An individual is treated as an individual in the testing interpretation. The student is referenced toward the behavior to be measured, rather than pre-established group norm performance. In effect, CRM affords individualized evaluation for individualized instruction. Second, the focus of CRM is on the subject content to be mastered, not on a vast array of numerical subscore continuums which mean little in and of themselves. Simple go-no go decisions are made on the basis of some mastery level that is stipulated for competency in the subject area, and concepts learned and those not yet mastered can be described in a simple fashion to educators and parents alike (Millman, 1970). Third, to build adequate measuring devices, and in turn to relate them back to the instructional process, the educator is forced to plan exactly what is to be taught by stating very specific behavioral objectives. Fourth, the fact that CRM mastery levels must be set forces the teacher to examine more intensively just what is and is not essential for continuous educational progress, since her go-no go decisions based upon CRM data affect directly the child's movement along the instructional continuum. Additional advantages could be listed. Because they are found so often in the literature, no formal review will be attempted here (see, for example, Bolvin & Glaser, 1968; Cox & Sterrett, 1970; Gorth, Grayson, Popejoy & Stroud, 1969; Johnson & Kress, 1971; Popham, 1970; Shoemaker, 1971;

Simon, 1969).

This paper will attempt to define what are believed to be distinguishing features of handicapped children which demand specialized measurement systems, not merely quick revampings of existing normal-child measurement technology. ("Handicapped" in this paper is to be interpreted as limited or impaired mental functioning.) A CRM system designed especially for the handicapped will be described. Finally, some CRM issues specific to the handicapped are listed. It is the contention of this paper that all too often the efforts of measurement specialists get directed toward the solution of problems for normal children, when in reality more generalizable schemes for all children could be produced if constraints for mentally handicapped children were kept in mind.

FACTORS THAT HAVE LED TO ANTIQUATED MEASUREMENT PROCEDURES FOR THE HANDICAPPED

All of the four major advantages of CRM described above have applicability to any "type" of child, whether he be normal, mentally retarded, emotionally disturbed, or whatever. However, there are some additional considerations peculiar to handicapped children that measurement experts rarely pay heed to. This situation is brought about not because measurement people are incapable of adapting their tools to handicapped populations, but rather because the whole area of "special" or "exceptional" education for the impaired child is neglected by most educational researchers. As a result, the many areas of impairment under the broad umbrella of "education for the handicapped" suffer innovative implementation lags much longer than even with the normal education realm.

First, there is a widespread belief that any student monitoring for the handicapped can be adequately approached by wise selection of standardized tests to match in-house instructional objectives, proper level of content difficulty, etc. However, beyond the purpose of screening and classifying handicapped children as being deviant from normal populations, little value can be derived from standardized tests for populations, monitoring children in an on-going fashion. While broad objectives might be matched from the instructional program to those of a standardized test, they are never specific enough to ferret out the root problems of a handicapped child. It matters not whether the standardized test is an achievement device or a diagnostic one. The standardized tests simply are too global in nature to be of much use to remediation of the child. Tests interpreted in a CRM sense and built to reflect the very specific, local school system objectives are much more useful to the handicapped children specialist. In summary, the usual global objectives, which seem to suffice in locating trouble spots for normal children, simply are not specific enough to pinpoint difficulties of mentally handicapped children.

A second point in connection with the misuse of standardized tests in guaging academic progress of handicapped students is the use of norm-referenced or group-oriented interpretations. As stated above, NRM devices have their appropriate role in identifying a "static" type of deviancy from the normal. However, it is one thing to identify a child as being, say, mentally retarded, and quite another to judge his progress ("dynamic") in terms of NRM. One already knows the child is deviant, and he will gain little by showing that his progress is also deviant ("A rose is a rose"!). Rather, we need for the deviant child a mode of interpretation that references his performance

to the criterion tasks and competency levels (CRM).

One strong bias in special educators and regular educators alike when it comes to building any type of detailed (in-depth) measurement system is that there just is not much to pick up at all. So why bother with any measurement methodology? Most special educators simply cannot see the value of building formal monitoring or accountability systems for mentally handicapped pupils. This sorry attitude has led to almost total lack of accountability in special education programs (see Proger, 1971).

CONSTRAINTS UPON MEASUREMENT SYSTEMS FOR THE HANDICAPPED

In dealing with several measurement experts from the bailiwick of regular education, it seems next to impossible for them to see what educators of the handicapped consider are unique testing problems. In part, this lack of communication is caused by the special educators who often find that they themselves cannot define precisely what they think is "special" about measurement in the world of special education (handicapped children). Special educators try to get across to evaluators that individualization of instruction is of prime importance to education for the handicapped and that elaborate item-examined sampling systems or any type of global-objective assessment simply is not applicable. "But," say the evaluators, "individualization of instruction is a prime goal of all education, not just for the handicapped. The IPI and CAM monitoring systems have fit in beautifully with individualized instruction." And this is as far as the argument usually gets, with the special educators wondering what they missed and the evaluators thinking they have hit the nail right on the head.

Unfortunately, there is more to the issue than meets the eye. There are two main considerations here. To discuss these, consider the two most common class plans for the handicapped. First, there is the small-group, self-contained class. Usually having no more than ten children, the self-contained class handles on a full-time, year-round basis those handicapped children who cannot keep up with their regular education counterparts. Second, there is the resource room which accommodates less severely handicapped children on a part-time basis; most of their time is spent in a regular class with only specialized individualization given in the resource room on a one-to-one basis.

In both classes, the two primary distinguishing features of instruction that directly affect measurement systems are (a) the individualization requires a more intensive subject-content-diagnostic in regular education orientation than (b) the individualization process not only allows different rates of progress, but also demands the flexibility whereby everybody in the one self-contained class or resource room might be on a totally different instructional approach (this is rarely ever the case in so-called individualized systems for normal children).

With respect to the first distinguishing feature, special educators usually rely upon very precise diagnostic information. This type of data simply is not forthcoming with typical global assessment packages or sampling schemes. Special educators of the mentally handicapped need very specific diagnostic information not even given by so-called diagnostic tests in reading. For example, consider the letter "a". It is one thing to learn whether a retarded child can visually discriminate the "a" from configurationally similar distractors, or whether he can perform a similar discrimination task auditorily. One can usually get this type of data from various diagnostic reading

tests. However, it is quite another thing to worry about whether the child can deal with "a" embedded in words, such as in the initial, medial, and final positions; whether the handicapped child might be better able to deal with "a" in visual modes of communication than auditory; whether the child has problems in copying or mimicking "a"; whether the child can deal effectively with the difference between long and short "a" and all exceptions thereto; and so on. These types of data cannot be obtained from the usual diagnostic reading tests; CRM is one vehicle appropriate to assessing such basic skills.

With regard to the second feature, in a regular classroom, even if individualization is practiced as well as preached, it is doubtful whether there will ever not be a common core of instructional sequence regardless of the pacing of students. Thus, for regular education, sampling schemes or just assessment of very global objectives does seem appropriate. However, there is just no way these testing approaches can accommodate a totally different instructional package for every child in a single class.

Besides worrying about the nature of instruction with handicapped children the testing process itself poses some constraints. The mentally handicapped child is hindered to a much greater extent by sensory processing difficulties than are normals. That is, one child might function very poorly in the visual modality and yet quite intactly with regard to the auditory one. Thus,one definitely wants to worry about assessing at least auditory and visual inputs on any task the child is given. This alone makes the testing game complicated enough. While it goes without saying that individualized test administration must be used (group tests are virtually meaningless with such groups as the retarded), this is not enough. Because of the

processing problems in the central nervous system, the perceptual problems, motor impairments, and so on, paper-and-pencil responding is out of the question for a lot of handicapped children. Fine motor coordination of any type (writing, darkening blank spaces, etc.) is usually a problem. Thus, pointing responses, vocal replies, etc., are perhaps more appropriate. Again, this can be done only in an individualized testing situation.

THE INDIVIDUAL ACHIEVEMENT MONITORING SYSTEM (IAMS) FOR HANDICAPPED CHILDREN

A CRM system for the handicapped was developed jointly under the auspices of PRISE (Pennsylvania Resources and Information Center for Special Education) and NRRC/P (National Regional Resource Center of Pennsylvania), for the NRRC/P project itself. NRRC/P is funded under Title VI - C of the Elementary and Secondary Education Act of 1965. The NRRC/P project has four divisions: Urban (Philadelphia), Middle Urban (Harrisburg and environs), Suburban (Philadelphia suburbs), and Rural (area around University Park, Pa. -- home of Pennsylvania State University). The CRM monitoring system, called the Individual Achievement Monitoring System (IAMS), is being developed and operated in the Suburban Division.[3]

For the advantages usually cited for CRM, as well as the special benefits for handicapped children even moreso than for normals, the entire accountability system was centered around CRM. Standardized tests are still used in the classic program evaluation designs, but these aspects are of minor interest to project personnel. In deciding to opt for a CRM accountability system, the first step was to examine currently functioning CRM-oriented programs that might possibly be adapted to the handicapped children's needs of NRRC/P.

The two main existing systems of CRM-oriented evaluation/instruction that were considered feasible were Individually Prescribed Instruction (IPI) and Comprehensive Achievement Monitoring (CAM).

Because IPI requires adherence to certain guidelines on the part of any school agency considering implementing the system (evidently to comply with their own research data-gathering needs), and because NRRC/P needed a great deal of flexibility in trying to make necessary modifications in a CRM network for the handicapped, IPI was ruled out. Turning toward CAM, NRRC/P personnel could not see how an item-examinee sampling framework could yield detailed enough pictures of <u>individual</u> children to make highly personalized programming decisions. In several classes for the handicapped within the umbrella of the NRRC/P auspices, group instruction is so rarely used that any type of <u>group</u> evaluations are meaningless. While many other factors entered into the decision not to adapt either CAM or IPI, it should be clear why the decision was made to build a CRM system using curriculum and evaluation expertise found within NRRC/P. Thus, "Phase One" of the IAMS dealt with getting a CRM system into practice as soon as possible for the 1971-1972 academic year.

While NRRC/P personnel found that a total existing CRM system could not be adapted readily to the project, an effort was still made to economize time, effort, and manpower by trying to adapt not entire systems, but components of systems. One crucial component is a coherent system of behavioral objectives. Thus, it was thought feasible to make the rounds of existing collections of behavioral objectives. Unfortunately, the classification systems of these objectives were simply too gross to allow specific individualized prescriptions and the appropriate CRM testing to go along with it. With all the talk and funded projects that have dealt with behavioral objectives, this

inability to tie in with a total systematic effort was quite frustrating. Not only were the collections of objectives not written within a complete classification framework, but also the gaps here and there at various levels made any linkage with NRRC/P unfeasible. A second component that NRRC/P thought it might be able to adapt in lieu of a total existing CRM system was a series of test items to match the very specific objectives. Several objectives collections had test items appended to the objectives. Again, the items simply did not cover specific enough behavior. Also, the auditory-visual processing problems mentioned earlier would not be accommodated. The general procedure of trying to adapt existing project components to form a new total instructional/evaluation system just was not feasible. There was also the occasional outbreak of "project paranoia" whereby people are reluctant to release materials "not quite completed," "in revision," and so on.

Thus, sadly enough, Phase One of the IAMS effort required building on its own a bank of objectives and test items to mesh with the type of diagnostic individualization peculiar to education of the mentally handicapped. In NRRC/P, children whose functioning levels of reading and arithmatic fall within the K to 6 range are admitted to its classes. The mentally handicapped children usually show some severe deficits in reading and/or arithmetic, and thus all instructional and remedial efforts in NRRC/P classes focus upon these two content areas. To simplify discussion, only the reading section of the IAMS will be described. To begin the task of building a very specific, coherent system of behavioral objectives, a linguistic- phonics program (Glim, 1968) was selected to serve as a model for a realistic instructional sequence that any teacher could tie into diagnostically. For purposes of diagnostic teaching,

objectives were written at the most specific level possible (e.g., "The child will mimic the /m/ sound presented orally," or, "The child will distinguish words with /m/ as the initial sound from words with other sounds in initial position presented orally.") Since a large number of handicapped children who have problems with reading must be helped at the basic skills level, these specific objectives are exactly what are needed. For each of these objectives so generated, multiple-choice auditorily-oriented or visually-oriented test items were written.

The CRM-guided instructional system was geared to three types of testing: placement, immediate achievement, and retention. The immediate achievement tests examine <u>every</u> specific objective of two-week chunks of instruction. That is, the curriculum sequence was divided ahead of time by curriculum experts into what they thought would be two-week periods of instruction. Thus, if a child simply is not getting much from the instructional process, he is discovered early enough before his learning problems multiply irreparably. A test, or monitor, in the IAMS system may run in length from eighty to about one hundred fifty items. Because of the intensity of this monitoring, the achievement monitors serve as measures of both global attainment and, in the case of failure, of diagnostic assessment. For retention, four two-week units of instruction were combined into an eight-week retention module, where sampling of objective-item mappings was employed. The length of a retention monitor,after sampling has been employed, reduces to that of a single achievement monitor. For the time being, the retention tests also serve as placement tests, the third category of IAMS testing (standardized diagnostic and achievement tests are also used to aid in placement decisions).

For purposes of intra- and inter-project reporting, the very specific objectives described above are combined into what is akin to course objectives (e.g., "The ability to identify, name, reproduce, and use in context the letter /s/ in all its forms, and in various positions in a given word.") There are usually from three to five of these course objectives for a unit of instruction. Percentage mastery scores are computed immediately after each testing by the teachers themselves. Feedback is therefore put to use when it can still make a difference. A criterion mastery level of 80% is set for <u>total</u> monitor achievement. In general, if a child reaches criterion on the specified body of material (two weeks' worth), he proceeds to the next instructional module or unit. If the child fails to reach criterion, he is put into an instructional branching network of either additional instruction (parallel branching) or remediation (backwards branching). The decision-making process with regard to instructional programming is illustrated in Tables 1, 2, and 3.

Insert Tables 1,2,3 about here

Thus far in this paper, only Phase One of NRRC/P's IAMS testing system has been considered. That is, the time-pressed procedures for gaining an immediately operational CRM system consumed all of the curricular and evaluation departments' time during the 1971-1972 academic year. To accomplish this end, a particular curricular sequence was selected for reading and for arithmetic. These sequences were task-analyzed in detail for both behavioral objectives, and corresponding test items were constructed for the CRM system. The big drawback to this CRM construction procedure is that the <u>sequence</u> of objectives and the <u>sequence</u> of test items within the two-week achievement

monitors is curriculum dependent. Phase Two of the NRRC/P monitoring efforts. therefore, was to build a CRM system of objectives and test items readily adaptable to any instructional package. Phase Two requires that test items and behavioral objectives be able to be rearranged in any sequence whatsoever.

To implement Phase Two, old "system-in-a-shoe-box" play was used. Every classroom will be provided with a large file drawer of behavioral objectives, test items for CRM construction, and resource ideas (worksheets, instructional materials, etc.) to implement the specific objectives. On one sheet of paper in the file system, the main mode of entry will be the behavioral objective. There will be a manual accompanying the file system to simplify the accession process. Once the teacher knows what specific objectives she wishes to work on in the near future, she will pull the appropriate sheets and arrange them in the instructional sequence she sees fit. (The primary advantage to Phase Two's CRM monitoring system is that the sequences of objectives and items can be rearranged to suit whatever needs the teacher has; this was not true of Phase One's CRM system.) Finally, if instructional resources are available (apart from trade books, readers, etc.), these are mentioned on the separate objective sheets.

Phase Two will be a refinement of the existing Phase One CRM system. Because of the nature of NRRC/P's handicapped children, a range of objectives covering K to 6 is sufficient for most needs. Phase One has dealt with K to 3. Thus, converting to Phase Two at the end of the 1971-1972 school year will require not only reworking the existing Phase One materials but also extending them upwards through the 4 to 6 range. The job will be greatly simplified by being able to obtain reading objectives for K to 6 from New York State

Education Department's SPPED project (System for Pupil and Program Evaluation and Development, directed by Robert P. O'Reilly). It should be noted that of the many objectives-oriented projects, SPPED was the only one that began to meet the needs of NRRC/P for handicapped children. (It was only after Phase One had been begun that SPPED's availability and appropriateness became known.

GENERAL ISSUES FOR DISCUSSION ON THE TOPIC OF CRM FOR THE HANDICAPPED

A perennial thought-provoker in CRM discussions is the mastery level, pass-fail, cutoff point issue. For normal children, there appear to be two general routes that one can travel. First, one can set an overall mastery level of X%, as IPI has been doing. The child either attains at least criterion, or he does not. Go-No Go. The Phase One IAMS of NRRC/P follows this procedure, setting 80% as the mastery level for all children to attain on all units of instruction. A second major cutoff point procedure is that of still requiring all children to attain the same mastery level on a given unit, but to vary the mastery levels from unit to unit, depending on the difficulty of material, importance of the material for later successful performance, etc. This second mastery level determination process seems more reflective of reality but is certainly much more difficult to implement, let alone justify specific levels decided upon. The specific issue for the handicapped that should be raised in connection with the mastery level cutoff point is whether either of the above procedures is appropriate for any kind of mentally handicapped child. Because of their limited mental potential, can

such an exceptional child be asked to achieve the same criterion level of mastery as a normal child, especially for more difficult topics? It would seem more appropriate -- although of doubtful feasibility -- to set mastery levels for each child relative to his potential.

With the issue of reliability, Livingston (1972) talks about reliability calculated not about the norm or mean but rather about the criterion level one sets up. Another possible approach would be to give the criterion test (posttest) as a pretest, on which a reasonable amount of variability should still exist (if mastery instruction has been successful, all children should be at criterion so that variability becomes quite restricted). Using the estimate of reliability (internal consistency) gotten from the pretest administration of the mastery test, would there be any stunning disadvantages to the pretest-derived reliability estimate versus the Livingston estimate?

Earlier in this paper in the section on "Constraints ...," the need for very specific, diagnostic monitoring was mentioned. This constraint alone would shed some doubt on the appropriateness of assessment of objectives written on a global level, as well as sampling schemes. In particular, item-examinee sampling seems to be inappropriate to handicapped populations. In sampling, one assumes some semblance of normality. Yet, it is well-known that not only are distributions of mentally handicapped children skewed, but they are multi-modal (see, for example, Nelson, 1970). In other words, there are several different, unique distributions to the area of "retardation" (genetic retardation, "slowness" in normals, accidentally produced brain damage, etc.). In view of these considerations, is there really any justification for any content or examinee sampling patterns with the handicapped?

What role does time play in the monitoring of handicapped children's performance (academic, social or physical)? The National Regional Resource Center of Pennsylvania believes time of instruction is quite relevant (total time needed to teach a specific objective) but seriously doubts time (rate) of responding is pertinent to the legitimate evaluation of handicapped children's progress. For mentally limited children, power rather than speediness seems to be the main concern of test administration. A few specific questions might be raised about daily (or at least frequent) monitoring that utilizes rate of responding as a criterion (such as the "Precision Teaching" movement, which began in 1965 at the University of Kansas). First, while an individual teacher might be helped by recording rates and interpreting graphs of those rates in her own way, one must always be cautious in judging the validity of such results. With emotional behavior (such as dealt with in behavior modification), the teacher must worry about what times during the day the behavior in question is recorded; if the behavior is frequent during the day, then sampling appears to be the only feasible answer. With academic performance, the difficulty of content (such as arithmetic problems) could fluctuate markedly from day to day, thus distorting the graphs of rates of responding. Further, the teacher has to concern herself about the comparability of sampled time periods from day to day; all types of outside contaminating influences can affect changes in rates in any graphical presentation. A second major problem with rate recording involves the appropriateness of rates of responding in comparison to absolute levels of performance. A child's rate graph could appear to show improvement simply by making the same number of errors but in less time; this situation could easily occur in solving sets of arithmetic problems. Also, there is the philosophical issue of whether one should consider the improvement of a disabled child in terms of quickness, or absolute quality, or both.

Another weird paradox or dilemma in measurement of on-going academic progress of the handicapped is the circular reasoning used to try to get at sensory deficits/learning styles. Because of the cruciality processing difficulties play in the achievement of any handicapped child, some measurement specialists believe any concept (in arithmetic or reading, for sake of argument) should be measured in at least two different ways: auditorily-based or visually-based. In the psycholinguistic processing model of Osgood, as modified by Kirk, McCarthy, and Kirk (1968), the auditory or visual emphasis to a test item can be introduced in any one or more of three dimensions: reception (stimulus), association (thought processes), or expression (response). Thus, for any item, there are eight (2 X 2 X 2) ways to modify the chain of events involved in a response to be primarily auditory or visual. This, of course, is not feasible to do in terms of test length alone. However, there appears to be a fallacious circularity of measurement reasoning to any testing of the reception-association-expression sequence. Starting at the top of the hierarchy, expression (responding) requires that both reception of stimuli and association of them with prior knowledge has occurred. Thus, expression is intimately dependent upon successful functioning in the processes of reception and association. One can never really claim he has attained anything near an uncontaminated measure of expression. One might respond that he can get out of this dilemma by partialling out, in effect, the lower-level effects of reception and association. Yet, to get any measure at all of the latter two processes, a large degree of expression (responding) is required. How does one exit this measurement disaster?

Finally, with the handicapped, there are some interesting phenomena regarding internal consistency of a given test. Consider a two-way design with type of child (normal and handicapped) and complexity of test (unidimensional and multidimensional). Assuming internal consistency is a function of the population upon which the data is to be generated, one could hypothesize for unidimensional tasks greater internal consistency for normals with intact processing mechanisms than for handicapped children. Even for highly homogeneous activities, for the handicapped these appear to be "different" as though they are not related at all. However, when one considers clearly multi-dimensional tests, one would predict internal consistency to change markedly (i.e., lower) for normals but negligibly for the processing-impaired or mentally handicapped child, who already sees unidimensional tests as multi-dimensional in effect.

SUMMARY

This paper has discussed criterion-referenced measurement (CRM) from a perspective other than that usually employed: handicapped children in distinction to normal children. It has been argued that all too often measurement experts try to generalize methodology used with normal children to handicapped children. Further, it has been argued that the measurement needs of programs for the handicapped are quite different than those for normal children. A CRM system devised especially for the handicapped is described: the Individual Achievement Monitoring System (IAMS). Finally, some CRM issues specific to the handicapped are discussed. It is the hope of this paper that measurement experts will take more than a cursory look at various types of handicapped populations with a view toward devising more appropriate measurement systems.

FOOTNOTES

[1]The writing of this paper was supported by PRISE under the Elementary and Secondary Education Act of 1965, Title III (Grant No. R-22-H, 48-70-0003-0). However, no endorsement of the content herein or the practices implied thereby is to be inferred on the part of PRISE or its funding source, the United States Office of Education.

[2]The authors would appreciate any criticisms, comments, or related views from others working in the CRM area. Please address all correspondence to Dr. Barton B. Proger, Director of Evaluation and Dissemination, Pennsylvania Resources and Information Center for Special Education, 443 South Gulph Road, King of Prussia, Pennsylvania 19406.

[3]Those desiring further information on aspects of NRRC/P other than the IAMS, (see footnote 2), should write to Dr. David L. Hayden, Director, National Regional Resource Center of Pennsylvania, Department of Special Education, Box 911, Harrisburg, Pennsylvania 17126.

REFERENCES

Bolvin, John O., and Glaser, Robert. Developmental Aspects of Individually Prescribed Instruction. Audiovisual Instruction, 1968, 13(8), 828-831.

Cox, Richard C., and Sterrett, Barbara G. A Model for Increasing the Meaning of Standardized Test Scores. Journal of Educational Measurement, 1970, 7, 227-228.

Glim, Theodore E. The Palo Alto Reading Program: Sequential Steps in Reading. New York: Harcourt, Brace & World, Inc., 1968.

Gorth, William P., Grayson, A., Popejoy, L., and Stroud, T. A Tape-Based Data Bank from Educational Research on Instructional Testing Using Longitudinal Item Sampling. Educational and Psychological Measurement, 1969, 29, 175-177.

Johnson, Marjorie Seddon, and Kress, Roy A. Task Analysis for Criterion-Referenced Tests. The Reading Teacher, 1971, 24, 355-359.

Kirk, Samuel A., McCarthy, James J., and Kirk, Winifred D. Illinois Test of Psycholinguistic Abilities (Revised Edition). Urbana, Ill.: University of Illinois Press, 1968.

Livingston, Samuel A. Criterion-Referenced Applications of Classical Test Theory, Journal of Educational Measurement, 1972, 9 (1), 13-26.

Millman, Jason. Reporting Student Progress: A Case for a Criterion-Referenced Marking System. Phi Delta Kappan, 1970, 52(4), 226-230.

Nelson, Thomas L. The Case for Multidisciplinary Research Efforts in Mental Retardation. The Journal of Special Education, 1970, 4(2), 177-179.

Popham, W. James (Ed.). Criterion-Referenced Measurement: An Introduction. Englewood Cliffs, N. J., Educational Technology Publications, 1971.

Proger, Barton B. Program Evaluation: The Model-Building Game. Journal of Learning Disabilities, 1971, 4(6), 292-306.

Shoemaker, David M. Criterion-Referenced Measurement Revisited. Educational Technology, 1971, 11(3), 61-62.

Simon, George B. Comments on "Implications of Criterion-Referenced Measurement." Journal of Educational Measurement, 1969, 6(4), 259-260.

FIGURE 1

GENERAL TYPES OF DEVICES FOR INDIVIDUAL ACHIEVEMENT MONITORING AND GLOBAL PROGRAM EVALUATION

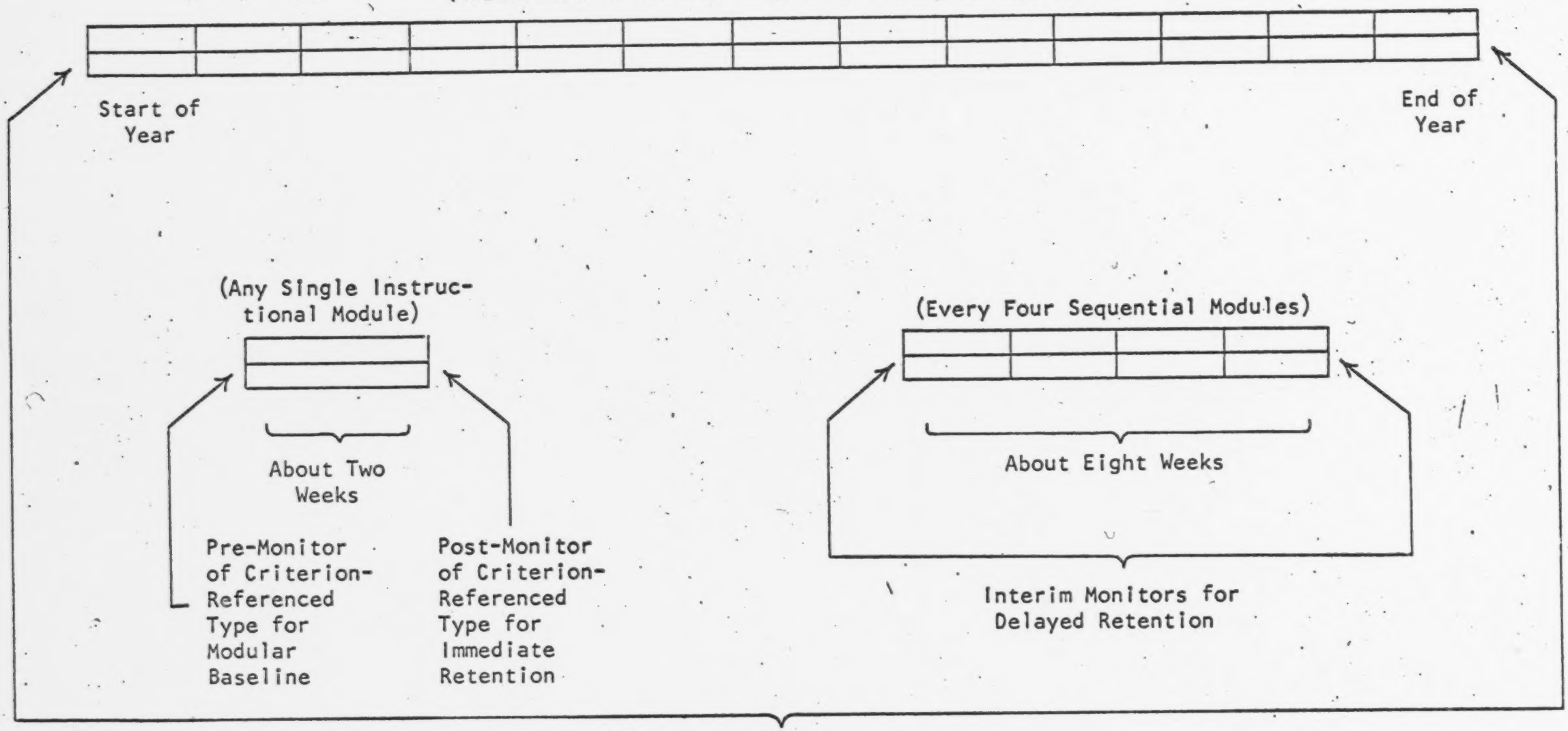

(Instructional Modules from Small-Group Mainstream)
Start of Year
End of Year
(Any Single Instructional Module)
About Two Weeks
Pre-Monitor of Criterion-Referenced Type for Modular Baseline
Post-Monitor of Criterion-Referenced Type for Immediate Retention
(Every Four Sequential Modules)
About Eight Weeks
Interim Monitors for Delayed Retention
Classical Standardized Tests Used at Year's Start for Screening and for Global Program Evaluation Pretest Levels and at Year's End for Global Posttest Levels

FIGURE 2

SMALL-GROUP MAINSTREAM OPERATIONAL SEQUENCE OF

INDIVIDUAL ACHIEVEMENT MONITORING SYSTEM

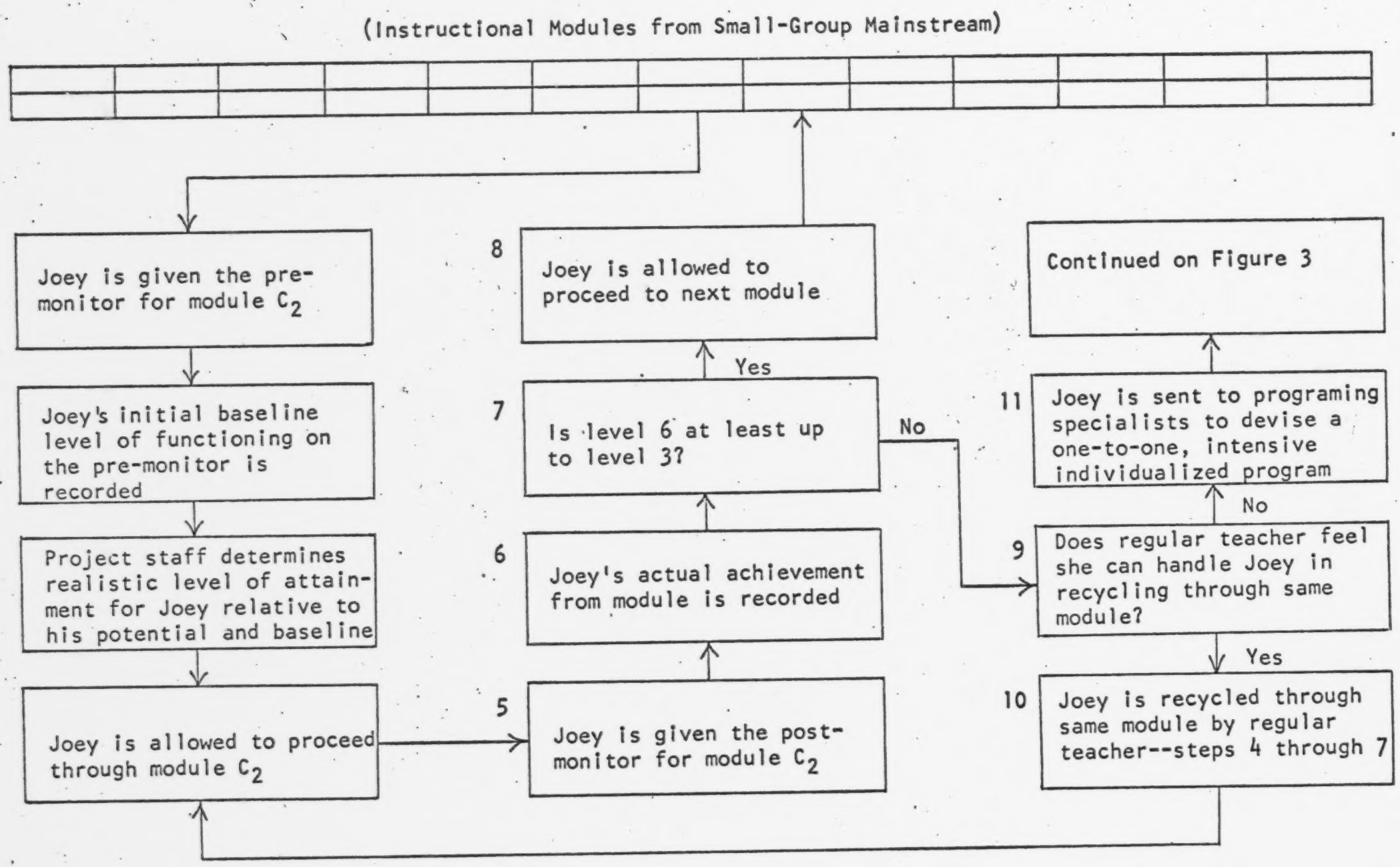
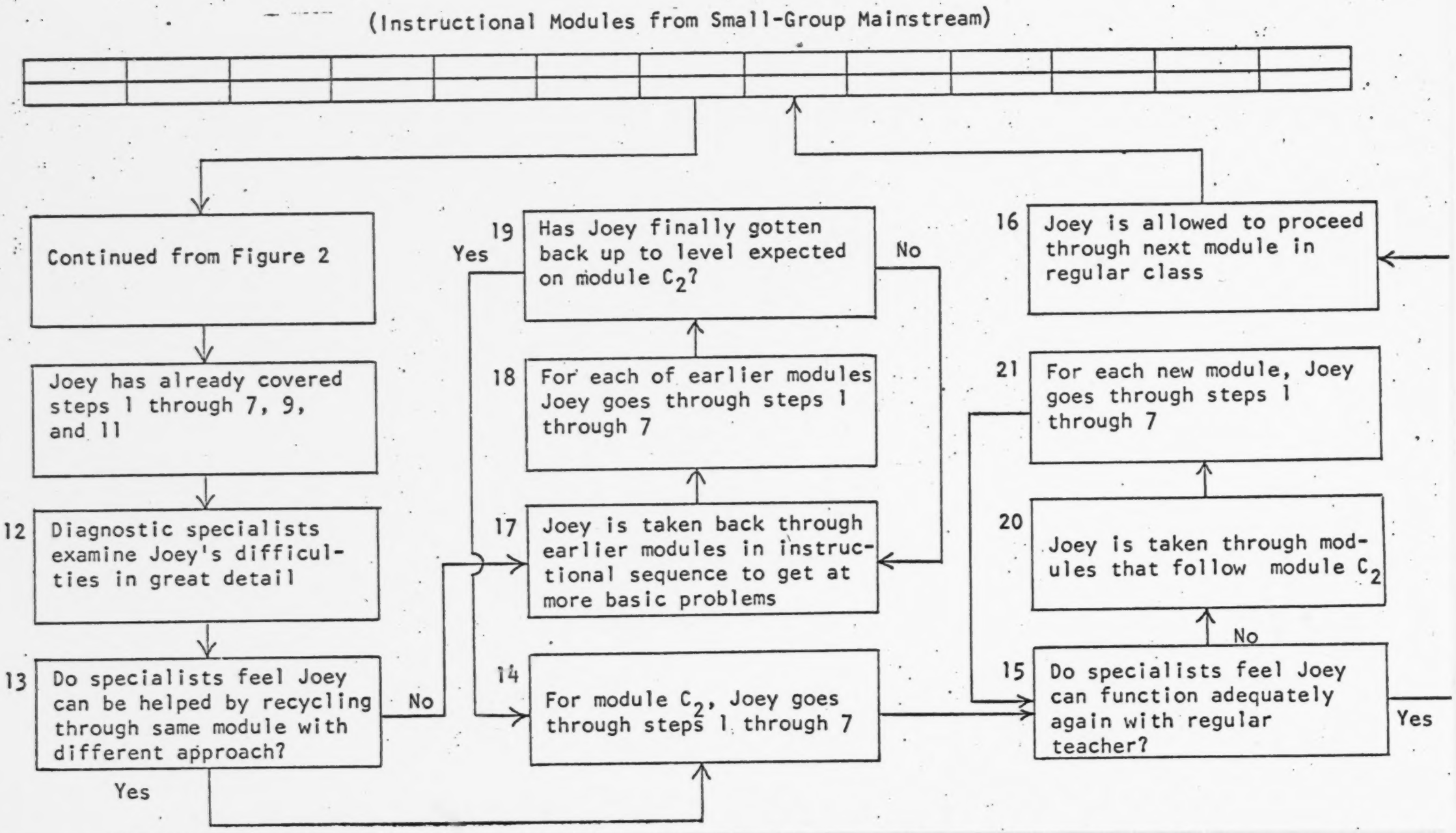

(Instructional Modules from Small-Group Mainstream)
1
Joey is given the pre-monitor for module C_2
2
Joey's initial baseline level of functioning on the pre-monitor is recorded
3
Project staff determines realistic level of attainment for Joey relative to his potential and baseline
4
Joey is allowed to proceed through module C_2
5
Joey is given the post-monitor for module C_2
6
Joey's actual achievement from module is recorded
7
Is level 6 at least up to level 3?
No
Yes
8
Joey is allowed to proceed to next module
9
Does regular teacher feel she can handle Joey in recycling through same module?
No
Yes
10
Joey is recycled through same module by regular teacher--steps 4 through 7
11
Joey is sent to programing specialists to devise a one-to-one, intensive individualized program
Continued on Figure 3

FIGURE 3

INSTRUCTIONAL BRANCHING POSSIBILITIES WITH SPECIALIZED PERSONNEL FOR CHILDREN WITH PARTICULARLY SEVERE PROBLEMS DURING INDIVIDUAL ACHIEVEMENT MONITORING


(Instructional Modules from Small-Group Mainstream)
Continued from Figure 2
Joey has already covered steps 1 through 7, 9, and 11
12 Diagnostic specialists examine Joey's difficulties in great detail
13 Do specialists feel Joey can be helped by recycling through same module with different approach?
No
Yes
19 Has Joey finally gotten back up to level expected on module C_2?
Yes
No
18 For each of earlier modules Joey goes through steps 1 through 7
17 Joey is taken back through earlier modules in instructional sequence to get at more basic problems
14 For module C_2, Joey goes through steps 1 through 7
16 Joey is allowed to proceed through next module in regular class
21 For each new module, Joey goes through steps 1 through 7
20 Joey is taken through modules that follow module C_2
No
15 Do specialists feel Joey can function adequately again with regular teacher?
Yes